

Experts' adjustment to model-based SKU-level forecasts: Does the forecast horizon matter?

Philip Hans Franses and Rianne Legerstee

*Econometric Institute
Erasmus School of Economics*

Abstract

Experts (managers) may have domain-specific knowledge that is not included in a statistical model and that can improve short-run and long-run forecasts of SKU-level sales data. While one-step-ahead forecasts address the conditional mean of the variable, model-based forecasts for longer horizons have a tendency to convert to the unconditional mean of a time series variable. Analyzing a large database concerning pharmaceutical sales forecasts for various products and adjusted by a range of experts, we examine whether the forecast horizon has an impact on what experts do and on how good they are once they adjust model-based forecasts. For this, we use regression-based methods and we obtain five innovative results. First, all horizons experience managerial intervention of forecasts. Second, the horizon that is most relevant to the managers shows greater overweighting of the expert adjustment. Third, for all horizons the expert adjusted forecasts have less accuracy than pure model-based forecasts, with distant horizons having the least deterioration. Fourth, when expert-adjusted forecasts are significantly better, they are best at those distant horizons. Fifth, when expert adjustment is down-weighted, expert forecast accuracy increases.

Key words: Model-based forecasts; expert forecasts; forecast accuracy; multi-step-ahead forecasts

This version: April 18 2010

Address for correspondence: Econometric Institute, Erasmus School of Economics, PO Box 1738, NL-3000 DR Rotterdam, The Netherlands, franses@ese.eur.nl, legerstee@ese.eur.nl. We thank the pharmaceutical company Organon BV (Oss, The Netherlands) for providing us with the data, and we thank two anonymous referees for their detailed comments.

1. Introduction

In supply chain management statistical models are regularly used to create out-of-sample forecasts for SKU-level sales data. One-step-ahead forecasts are useful for various reasons, but for management it may sometimes be more relevant to have forecasts for more distant horizons. The parameters in statistical models for time series data are typically estimated or updated using some least squares criterion, where usually the squares of the one-step-ahead forecasts are minimized. Hence, it is conceivable that statistical models would do well for short horizons but lesser so for more distant horizons. Even though statistical models seek to capture trends and other patterns in the data and to extrapolate these into the future, it is quite likely that these models do not cover recent events that can be relevant for forecasting too. Experts with domain-specific knowledge can have information on these events and hence can add something to the model-based forecast, see Fildes et al. (2009).

There is much evidence that managers adjust statistical model-based forecasts for SKU-level time series data (see Fildes and Goodwin (2007), Bunn and Salo (1996), Sanders and Manrodt (1994), Nikolopoulos et al. (2005), Syntetos et al. (2009), Fildes et al. (2008) and Syntetos et al. (2009)). Usually, SKU-level time series forecasts are created from extrapolation techniques like exponential smoothing and trend fitting, and this may motivate managers to adjust as they know of relevant factors other than the recent past of the sales data. Goodwin (2000) summarizes also other reasons to adjust statistical forecasts. There are studies that suggest that such managerial intervention leads to improved forecasts (Mathews and Diamantopoulos (1986) and Diamantopoulos and Mathews (1989)), but a recent account in Fildes and Goodwin (2007) suggests that model-based SKU-level sales forecasts may be adjusted too often, leading to a decrease in accuracy.

As the literature so far looked only at one-step-ahead forecasts, in this paper we are the first to study the behaviour of experts across forecast horizons. Given that model-based forecasts for stationary series for more distant horizons convert towards the unconditional mean of that variable, it might be that experts have a tendency to change those forecasts more than those for nearer forecast horizons. We shall study these and other questions using a large database with sales forecasts for pharmaceutical products.

The outline of our paper is as follows. To shape our empirical analysis we put forward a few research questions in Section 2. Next, we outline the empirical methodology in Section 3. Section 4 deals with the empirical results and Section 5 sketches a few areas for further work.

2. Research questions

This paper builds and extends on the insights in Fildes et al. (2009) and Franses and Legerstee (2009) where the question was addressed whether expert-adjusted forecasts have more accuracy. Among the many findings in the recent literature, it is documented that experts have a tendency to be over-optimistic, that expert forecasts show bias, but also that improvement can be noticed under various circumstances, see Fildes et al. (2009).

In the present study, and this is done for the first time in the literature, we consider not just the one-step-ahead forecasts but also multiple-step-ahead forecasts.

Variables of interest

The framework is as follows. We consider the following variables

$MF_{t+h t}$:	model-based forecast for horizon h (made from origin t)
$EF_{t+h t}$:	expert forecast for horizon h (made from origin t)
S_{t+h} :	realization at time $t + h$

where S denotes SKU-level sales, where our data to be analyzed below concern monthly data.

The model-based forecasts typically are linear functions of past sales, and in our illustration they are created by (a variant of) the ForecastPro program. The forecasts are created recursively which means that the parameters are estimated for R in-sample data, and then one-step-ahead to h -step-ahead forecasts are made. Next, the sample is enlarged to $R+1$, parameters are updated and again the h forecasts are made. When the total sample contains $R+P$ observations, there are thus P one-step-ahead forecasts to evaluate. Similarly, there are $P-1$ two-step-ahead forecasts, and likewise, there are $P-(h-1)$ h -step-ahead forecasts.

In supply chain management it is quite common that experts receive the statistical model-based forecasts and are allowed to modify these. This is of course partly due to the fact that the experts know that the model-based forecasts are based on a limited set of explanatory variables. It is however unknown to what extent the experts take account of the model-based forecasts. As Boulaksil and Franses (2009) document, about half of the surveyed experts claim to ignore the model-based forecasts when creating their own.

One way to express this notion is to assume that

$$(1) \quad EF_{t+h|t} = \lambda MF_{t+h|t} + A_{t+h|t}$$

When $\lambda = 0$, the expert fully ignores the model forecasts. In what follows, we define the contribution of the expert as $A_{t+h|t}$, which can be estimated by applying OLS to (1), see also Blattberg and Hoch (1990) and Franses and Legerstee (2009). It can happen for a particular moment t and horizon h that $EF_{t+h|t}$ equals $MF_{t+h|t}$ and in that case $A_{t+h|t}$ is equal to 0. We decide not to discard these data, as that would create missing observations in the time series of $A_{t+h|t}$ and we need the full time series in our model (2) below. We will use the same database as in Franses and Legerstee (2009), and in that paper it was already documented that expert adjustment occurs very frequently and this holds for all horizons (unreported to save space). It never happens that for all observations t $EF_{t+h|t}$ equals $MF_{t+h|t}$.

Research questions

The main issue in this paper is whether the forecast horizon matters for the behaviour of experts and for their forecast accuracy relative to model-based forecasts. We could find no relevant literature, so we stress that the following must be seen as exploratory analysis.

The first question is that we wish to examine is whether the added contribution of the expert is about equal for all forecast horizons. This implies that model-based forecasts for all horizons experience adjustment, and that there is no particular reason to believe that some horizons see lesser inclination by experts to adjust.

Whether experts are equally successful across horizons is another issue of interest. In fact, one might expect that experts may have access to recent information that never could have been included in the model, so the quality of the added contribution of the expert could be better for very short forecast horizons. On the other hand, regression-based models have a tendency to give h-step-ahead forecasts for a stationary time series that converge to the unconditional mean of that series (and also exponential smoothing does that). Hence, as such models may miss recently established trends that could be noticed by the expert, the quality of the added value of the expert may be larger for more distant horizons. In sum we therefore investigate whether the quality of the added contribution of the expert is larger for immediate and more distant horizons, and is smaller for intermediate horizons. In our empirical analysis below we will label 1, 2 and 3 months ahead as short-term, 10, 11 and 12 months ahead as long-term and around 4 to 9 months as mid-term horizons.

Based on our discussions with the managers at the headquarters' office, from which we draw our data and where they create the model-based forecasts, we know that the 6-month-ahead forecasts are most relevant for their supply chain management. So, for these horizons the experts do want to perform particularly well, and in fact part of their bonus payments depends on their success rates for that horizon. Given this information, we specifically focus on this mid-term horizon.

Finally, Franses and Legerstee (2010) find that when the weight of the contribution of the expert relative to the model is larger, the quality of the expert forecast becomes smaller. Their study addresses only one-step-ahead forecasts, and here we examine whether their findings carries through for multiple-step-ahead forecasts.

3. Methodology

This section gives a discussion of the database we have and it outlines the basic statistics that we compute to examine the hypotheses. We have data concerning product i within category j for country c . We consider 37 countries and there are 7 product categories. We do not have all data for all categories for all countries and within a category there are different numbers of products. The data concern monthly SKU-level sales for October 2004 to October 2006 of pharmaceutical products. The headquarters' office creates model-based forecasts and sends these to the experts in each of the countries. Each country has one or two experts and they are allowed to modify the model-based forecasts in a way they see fit. The database is very large and to make further analysis tractable, we decide to use the country-category unit as the unit of measurement. Depending on the availability of the h -step-ahead forecasts, we have about 180 to 220 cases for each of the horizons. In the Appendix we provide a few basic statistics, which show that expert forecasts usually differ from model forecasts and that the percentage change is substantial across all horizons.

The added value of the expert

First we need to create a measure for the contribution of the expert. There are various ways to do this, and we follow the same strategy as in Franses and Legerstee (2009). For this, we

consider for all products within a country-category combination the following multiple-equation model¹ for each horizon h , where we now consider $h = 1, 2, 3, 4, \dots, 11, 12$, that is,

$$(2) \quad \begin{aligned} S_{1,t+h} &= \alpha_1 + \beta MF_{1,t+h|t} + \gamma A_{1,t+h|t} + u_{1,t+h} \\ S_{2,t+h} &= \alpha_2 + \beta MF_{2,t+h|t} + \gamma A_{2,t+h|t} + u_{2,t+h} \\ &\dots \\ S_{n,t+h} &= \alpha_n + \beta MF_{n,t+h|t} + \gamma A_{n,t+h|t} + u_{n,t+h}, \end{aligned}$$

where $S_{i,t+h}$ denotes the SKU-level sales of the i -th product within a country-category unit, where $MF_{i,t+h|t}$ is the related model-based forecasts and where $A_{i,t+h|t}$ denotes the residual from (1), with $i = 1, 2, \dots, n$, with n the amount of products within each such combination. For some country-category units n is as small as 1, for others it can be 30. The parameters of interest are β and γ , which are pooled across the n equations to gain degrees of freedom.² After examination of the time series properties of each of the errors $u_{i,t+h}$, we find that the error processes are best modeled as

$$(1 - \rho_1 L - \rho_2 L^2)u_{i,t+h} = \varepsilon_{i,t+h}$$

where $\varepsilon_{i,t+h}$ denotes a standard white noise process. Each of these processes has a different variance to accommodate different scales of $S_{i,t+h}$. Here we see that having a complete set of observations on $A_{t+h|t}$ is important for the analysis of (2).

We allow the α_i parameters in (2) to differ per product, but the β and γ parameters are assumed as constant across products within a country-category combination. The model in (2) assumes independent equations with cross-equation parameter restrictions as we assume the errors $\varepsilon_{i,t+h}$ as independent. The model parameters are estimated using ordinary least squares (OLS) with the cross equation restrictions imposed. The covariance matrix of $\varepsilon_{i,t+h}$ is a diagonal matrix.

¹ This model is also used in Franses and Legerstee (2009) for the case $h = 1$.

² Fildes et al. (2009) found that the parameter values could differ across positive and negative values of A . We postpone a detailed analysis of this issue to further work.

For each of the country-category units and for each horizon h , we get an estimate of β and of γ . We are interested in testing whether $\gamma = 0$ and if it is not, if $\beta = \gamma$, implying that the model forecast and the added expert knowledge have equal weights in explaining $S_{i,t+h}$. The relative value of γ versus that of β gives an indication as to how much weight the added expert adjustment gets.

The quality of the expert

To examine the quality of the added contribution of the expert, one can of course just see if the root mean squared prediction errors (RMSPEs) or mean absolute percentage errors (MAPEs) are different, but one may also want to test whether the differences are significantly different. Such a test can then be computed for each product within each country-category combination.

To test the null hypothesis that the RMSPE of the expert forecast is equal to that of the model against the alternative hypothesis that the expert forecast is better, we need to take account of the fact that expert forecast potentially nests the model used for the statistical forecast, see (1) and Franses and Legerstee (2010) for details. For MAPE such a test does not exist. We thus follow the recommendation in Clark and McCracken (2001) who recommend the so-called ENC_NEW test defined by

$$(3) \quad ENC_NEW_h = P \frac{\frac{1}{P} \sum (e_{1,t+h|t}^2 - e_{1,t+h|t} e_{2,t+h|t})}{\frac{1}{P} \sum e_{2,t+h|t}^2}$$

The summation runs for the P one-step-ahead forecasts, and $e_{1,t+h|t}$ denotes the errors for model-based forecasts, and $e_{2,t+h|t}$ concerns the expert forecasts. Following Franses and Legerstee (2010), for one-step-ahead forecast errors we set the 5% critical value at 1.481. The test in (3) is a one-sided test of the null hypothesis that the expert forecast is equally good as the model-based forecast against the alternative hypothesis that the expert forecast is better.

For multi-steps-ahead forecasts we can also compute the test statistic as in (3), but now a complication arises in terms of its asymptotic distribution. For multiple-steps-ahead forecasts it is well known that the forecast errors are correlated, and this correlation needs to be included in the distribution. Clark and McCracken (2005) outline in detail how to do this in

case of a linear regression model. These authors recommend a one-sided test with a 5% critical value equal to 1.645.

4. Results

In this section we report on the empirical results obtained using the methodology outlined in the previous section, where the specific focus is on the variation of the results across horizons.

4.1 Added contribution of experts

In total we have data for 203 country-category combinations concerning one-step-ahead forecasts, and within each we have from 1 to as many as 30 products, see Franses and Legerstee (2009, 2010) for more details on the database. Table 1 gives in its second column how many cases we have for each of the 12 different horizons. We observe that we have quite a number of cases, and hence, we can safely draw generalizing conclusions.

Insert Table 1 about here

The third column of Table 1 gives the number of cases with significant values for γ , meaning that the added contribution of expert adjustment is relevant. The fourth column shows that the fraction of cases with significant γ seems to decrease with the forecast horizon. However, we must bear in mind that the test for this significance is based on smaller samples each time the horizon increases. In fact, for forecast horizon $h = 1$ each equation in (2) gets estimated for 25 observations, while for horizon $h = 12$ it is only 14. To provide some form of correction for this power loss, we compute a sample-size-corrected fraction of cases with significant γ by multiplying the numbers in the penultimate column of Table 1 by

$$\sqrt{\frac{T}{T-h+1}}$$

where T equals 25 here. Upon doing so, we see from the last column that the fraction of cases with significant γ in (2) is rather constant across the forecast horizons, and it is around 50% of the cases. So, in about half of the cases, the contribution of the expert is helpful, and this

fraction is constant across horizons. In sum, all horizons experience managerial intervention of forecasts and this is relevant in a constant fraction of the cases.

Insert Table 2 about here

Now we look at the added contribution relative to that of the model using the multiple-equation model in (2). In Table 2 we report on the fraction of cases where β equals γ , first across all cases and second across only those cases where γ is significantly different from 0. Again, we correct for the decreasing sample size with increasing horizon, see column 5 of Table 2, but as it now concerns the size of the test we compute the sample-size adjusted fractions by

$$1 - \sqrt{\frac{T}{T-h+1}}(1 - \text{fraction})$$

When we consider the relevant columns in Table 2 we observe that there is a slow decrease in the fraction of cases with β equal to γ , and this certainly holds for the cases with significant γ parameters. Hence, from Table 1 we saw that the expert contribution is stable across horizons, and from Table 2 we see that the number of cases where the model and expert are equally important becomes smaller as the horizon increases.

Insert Table 3 about here

To have a closer look at these relative weights, we compute the mean and the median of the estimated parameters in (2), that is, $\beta - \gamma$ across the cases mentioned in the second column of Table 1 and for the cases where γ is significant (the cases in the third column of Table 1). The median values in left-hand panel of Table 3 are already suggestive, but most clear-cut evidence is obtained from the median values in the right-hand panel of Table 3. The parabolic shape that is visually obvious for these numbers is even further substantiated when we consider the regression

$$(4) \quad \text{Median}_{\beta-\gamma} = \mu + \alpha_1 h + \alpha_2 h^2 + \varepsilon$$

for $h = 1, 2, \dots, 12$. The α_1 parameter gets estimated as -0.071 (with standard error 0.020) and α_2 becomes 0.006 (0.001) and the R^2 is 0.602. The minimum value of the fit of (4) is attained at $h = 6$, which is indeed the most important forecast horizon for the managers in our illustration. Hence for this forecast horizon, the expert contribution relative to the model forecast is highest. So, our second finding is that the horizon that is most relevant to the managers shows greater overweighting of the expert adjustment

4.2 Quality of added contribution

Now we have seen that experts add most to the model-based forecasts for horizon 6, the next obvious question is whether this is successful or not.

Insert Table 4 about here

In Table 4 we present the mean and median values of the empirical distribution of the percentage of improvement in RMSPE and MAPE of the expert-adjusted forecasts over the model-based forecasts for all cases (as in Table 1, column 2). At first sight for RMSPE we again see a parabolic shape, with a dip around horizons 6 or 7. This is confirmed when we estimate the parameters of

$$(5) \quad \text{Median \%improvement in RMSPE} = \mu + \alpha_1 h + \alpha_2 h^2 + \varepsilon$$

for $h = 1, 2, \dots, 12$. The α_1 parameter gets estimated as -0.713 (with standard error 0.347) and α_2 becomes 0.060 (0.026) and the R^2 is 0.398. Again, the fit of this regression attains its minimum value when $h = 6$. For MAPE the same regression gives a minimum value around 3 to 4. We also see that for more distant horizons expert adjusted forecasts are less bad. Our third main conclusion is that what was found for one-step-ahead forecasts in Franses and Legerstee (2010) also holds for the other 11 horizons. Expert adjusted forecasts are less accurate than pure model-based forecasts. We also see that the distribution is skewed to the left meaning that very poor expert forecasts occur more often than very poor model forecasts. In sum, for all horizons the expert adjusted forecasts have less accuracy than pure model-based forecasts, with distant horizons having the least deterioration..

Insert Table 5 about here

The conclusion drawn from Table 4 was based on the analysis of all forecasts, and the next question is whether this finding would also hold when we only look at the significant positive contributions of the experts, which now cover the averages within categories for the products with significant test values. When we look at the percentage improvement of experts in case the ENC_NEW test values are significant, as in Table 5, we observe that when the expert has a significant and positive added value to the final forecast, this value becomes larger for more distant away horizons. So, when experts do better, they do better for more distant horizons.

Taking the results in the last tables and figures together, we see that when experts do significantly better they do even more so when the forecast horizon increases. Moreover, when they do not necessarily perform significantly better, their forecasts are worst around the horizon that is most important for their business. This matches with the findings in Fildes et al. (2009) and Franses and Legerstee (2009) that when experts exercise too much value in the final forecast, notably here around horizon 6, the quality of this forecast quickly deteriorates.

Insert Table 6 about here

Following the discussion in Franses and Legerstee (2010), to demonstrate that less added contribution of the expert leads to improvement for all horizons, we report in Table 6 on the mean and median of the percentage of improvement in RMSPE and MAPE when we would take the average of the model-based forecast and the expert forecast, versus only the model-based forecast. Clearly, the median is positive across all cases and the minimum values are not that low anymore. Most improvement is seen for horizons 1 and 12, and lesser improvement for $h = 6$ and $h = 7$.

5. Conclusions

We analyzed model-based forecasts and expert-adjusted forecasts for a large range of products within various categories and where the experts are located in countries spread all over the world. We arrive at the following four conclusions concerning the impact of the forecast horizon on relevance and quality of experts' added contribution to model-based forecasts. First, all horizons experience managerial intervention of forecasts. Second, the

horizon that is most relevant to the managers shows greater overweighting of the expert adjustment. Third, for all horizons the expert adjusted forecasts have less accuracy than pure model-based forecasts, with distant horizons having the least deterioration. Fourth, when expert-adjusted forecasts are significantly better, they are best at those distant horizons. Fifth, when expert adjustment is down-weighted, expert forecast accuracy increases.

To answer the question in the title of our paper: Yes, the forecast horizon matters. This finding has consequences for management, in particular when training the experts. It is seen that experts exercise too much impact when the horizons matters most to them and then the quality of the added expertise quickly decreases. Apparently, experts should be trained not to over-adjust model-based forecasts for some horizons.

Appendix: Summary statistics of the data

Horizon	Percentage of observations where Expert Forecast \neq Model Forecast	Average Absolute Percentage Difference Between Expert Forecast And Model Forecast
---------	---	--

=====

1	0.96	71.32
2	0.95	76.60
3	0.95	85.01
4	0.95	90.65
5	0.95	97.86
6	0.95	109.40
7	0.95	113.98
8	0.95	115.96
9	0.95	127.09
10	0.95	79.91
11	0.96	137.99
12	0.96	72.06

=====

Note: To compute the numbers in the last columns: when the model forecast at time t is 0, the forecast at time $t+1$ is taken

Table 1:

Testing for the relevance of the added value of the expert by testing whether γ is significant in the system of equations in (2)

Horizon	Cases	Cases with $\gamma \neq 0$	Fraction	
			No correction	Corrected for power loss
1	203	105	0.517	0.517
2	201	109	0.542	0.553
3	201	103	0.512	0.534
4	199	96	0.482	0.514
5	199	90	0.452	0.493
6	196	87	0.444	0.496
7	194	88	0.454	0.521
8	189	78	0.413	0.487
9	189	75	0.397	0.481
10	190	76	0.400	0.500
11	186	65	0.349	0.451
12	185	69	0.373	0.498

Table 2:

Is the added contribution of the expert equal to that of the model, that is, is β equal to γ in (2)?
 (Fractions are in parentheses, corrected for sample size)

Horizon	Cases	Cases with $\beta = \gamma$	Cases with $\beta = \gamma$ and $\gamma \neq 0$
1	203	124 (0.611) (0.611)	54 of 105 (0.514)
2	201	118 (0.587) (0.578)	53 of 109 (0.486)
3	201	129 (0.642) (0.627)	53 of 103 (0.515)
4	199	126 (0.633) (0.609)	47 of 96 (0.490)
5	199	125 (0.628) (0.594)	41 of 90 (0.456)
6	196	119 (0.607) (0.561)	36 of 87 (0.414)
7	194	125 (0.644) (0.592)	40 of 88 (0.455)
8	189	123 (0.651) (0.589)	40 of 78 (0.513)
9	189	118 (0.624) (0.544)	33 of 75 (0.440)
10	190	111 (0.584) (0.480)	36 of 76 (0.474)
11	186	106 (0.570) (0.445)	22 of 65 (0.338)
12	185	111 (0.600) (0.465)	21 of 69 (0.304)

Table 3:

Empirical distribution of the estimated value of $\beta-\gamma$ in (2) for all cases (number of cases is given in Table 1, second column) and for the cases where $\gamma \neq 0$ (number of cases is given in Table 1, third column)

Horizon	All cases		Cases with $\gamma \neq 0$	
	Mean	Median	Mean	Median
1	0.082	0.074	0.064	0.041
2	-0.015	0.044	-0.158	-0.113
3	-0.001	-0.020	-0.193	-0.156
4	0.080	0.039	-0.133	-0.067
5	0.029	0.101	-0.091	-0.128
6	0.037	-0.006	-0.182	-0.235
7	-0.019	0.002	-0.165	-0.161
8	0.143	0.056	-0.135	-0.166
9	0.075	0.066	-0.093	-0.098
10	0.214	0.102	0.062	-0.114
11	0.185	0.134	0.163	0.015
12	-0.093	0.077	-0.148	-0.049

Table 4:

Empirical distribution of the percentage of improvement in RMSPE and MAPE of expert-adjusted forecast over the model-based forecast (all cases given in Table 1)

Horizon	RMSPE		MAPE	
	Mean	Median	Mean	Median
1	-19.892	-2.294	-32.304	-7.610
2	-23.950	-2.599	-56.888	-10.145
3	-26.747	-3.129	-35.394	-9.234
4	-30.641	-2.840	-34.738	-6.758
5	-25.799	-2.850	-33.870	-8.655
6	-24.911	-3.045	-32.226	-6.051
7	-23.309	-3.735	-32.777	-8.934
8	-17.998	-4.740	-38.615	-7.776
9	-17.121	-1.580	-24.418	-5.562
10	-304.43	-2.760	-96.575	-4.222
11	-18.289	-3.347	-22.832	-5.531
12	-12.357	-0.230	-16.387	-0.996

Table 5:

Distribution of the percentage of improvement in RMSPE and MAPE of expert-adjusted forecasts over the model-based forecasts, in case the ENC-NEW test is positive and significant and improvement is not zero (test is performed for each product in all categories and the cell only concern those products where the test is significant)

Horizon	RMSPE		MAPE	
	Mean	Median	Mean	Median
1	17.266	15.480	22.802	19.642
2	17.419	15.400	20.958	19.129
3	17.752	15.120	21.059	19.082
4	18.382	15.870	21.433	19.438
5	20.401	17.460	23.606	21.886
6	20.413	17.795	24.202	21.000
7	19.943	17.390	25.075	22.598
8	21.840	18.180	26.063	24.413
9	21.022	18.330	26.115	24.242
10	21.802	19.510	27.451	24.117
11	21.989	20.060	26.469	22.915
12	24.412	22.515	28.992	28.553

Table 6:

Empirical distribution of the percentage of improvement in RMSPE of expert-adjusted forecast over the model-based forecast, if the final forecast is 50-50 (all products in all cases given in Table 1)

Horizon	RMSPE		MAPE	
	Mean	Median	Mean	Median
1	1.571	7.010	-4.736	3.045
2	-0.113	6.250	-4.743	2.009
3	-1.229	6.400	-5.173	3.044
4	-3.318	6.790	-5.574	2.832
5	-0.577	6.360	-4.509	2.960
6	-0.240	6.220	-4.445	3.639
7	-0.825	5.965	-5.957	3.030
8	2.791	6.130	-1.480	3.115
9	3.175	7.090	-0.901	4.109
10	-140.27	5.915	-37.82	4.960
11	2.867	6.440	-0.311	5.429
12	4.965	8.490	2.036	7.177

References

- Blattberg, Robert C. and Stephen J. Hoch (1990), Database models and managerial intuition: 50% model + 50% manager, *Management Science*, 36, 887-899
- Boulaksil, Y, and P.H. Franses (2009), Experts' stated behavior, *Interfaces*, 39, 168-171
- Bunn, D.W. and A.A. Salo (1996), Adjustment of forecasts with model consistent expectations, *International Journal of Forecasting*, 12, 163-170
- Clark, Todd E. and Michael W. McCracken (2001), Tests of equal forecast accuracy and encompassing for nested models, *Journal of Econometrics*, 105, 85-110
- Clark, Todd E. and Michael W. McCracken (2005), Evaluating direct multi-step forecasts, *Econometric Reviews*, 24, 369-404
- Diamantopoulos, A. and B.P. Mathews (1989), Factors affecting the nature and effectiveness of subjective revision in sales forecasting: An empirical study, *Managerial and Decision Economics*, 10, 51-59
- Fildes, R. and P. Goodwin (2007), Good and bad judgement in forecasting: Lessons from four companies, *Foresight*, Fall 2007, 5-10
- Fildes, R., P. Goodwin, M. Lawrence and K. Nikolopoulos (2009), Effective forecasting and judgemental adjustments: An empirical evaluation and strategies for improvement in supply-chain forecasting, *International Journal of Forecasting*, 25, 3-23
- Fildes, R., K. Nikolopoulos, S.F. Crone and A.A. Syntetos (2008), Forecasting and operational research: A review, *Journal of the Operational Research Society*, 59, 1150-1172.
- Franses, P.H. and R. Legerstee (2009), Properties of expert adjustments on model-based SKU-level forecasts, *International Journal of Forecasting*, 25, 35-47

Franses, P.H. and R. Legerstee (2010), Does experts' adjustment to model-based SKU-level forecasts improve forecast quality?, *Journal of Forecasting*, 29, 331-340.

Goodwin, P. (2000), Improving the voluntary integration of statistical forecasts and judgement, *International Journal of Forecasting*, 16, 85-99

Mathews, B.P. and A. Diamantopoulos (1986), Managerial intervention in forecasting: An empirical investigation of forecast manipulation, *International Journal of Research in Marketing*, 3, 3-10

Nikolopoulos, K., R. Fildes, P. Goodwin and M. Lawrence (2005), On the accuracy of judgemental interventions on forecasting support systems, Lancaster University Management School Working paper 2005/022

Sanders, N. and K.B. Manrodt (1994), Forecasting practices in US corporations: Survey results, *Interfaces*, 24, 92-100

Syntetos, A.A., J.E. Boylan, and S.M. Disney (2009), Forecasting for inventory planning: A 50 year review, *Journal of the Operational Research Society*, 60, 5149-5160.

Syntetos, A.A., K. Nikolopoulos, J.E. Boylan, R. Fildes and P. Goodwin (2009), The effects of integrating management judgement into intermittent demand forecasts, *International Journal of Production Economics*, 118, 72-81.